• Technical Colloquium: Quantifying the Weight of Forensic Evidence
• May 5, 2016, NIST, Gaithersburg, MD
WoE in the Three Fields

How lawyers speak about WoE

How statisticians speak about WoE

How forensic scientists speak about WoE
Whether the spill or the initially mislabeled autoradiograph affected the reliability of the test is a question of fact. Alleged infirmities in the performance of a test usually go to the weight of the evidence, not to its admissibility. ... [T]he irregularities which occurred here do not warrant ... exclusion.

Weight vs Relevance

Evidence is relevant if:

(a) it has any tendency to make a fact more or less probable than it would be without the evidence; and

(b) the fact is of consequence in determining the action.

\[ P(H|E) \neq P(H) \]

- Fed R Evid 401
Weight = Probative Value

The court may exclude relevant evidence if its probative value is substantially outweighed by a danger of one or more of the following: unfair prejudice, confusing the issues, misleading the jury, undue delay, wasting time, or needlessly presenting cumulative evidence.

Fed R Evid 403
Weight vs Sufficiency

But your honor, the verdict is against the weight of the evidence!

E.g., Pa. R. Crim. P. 607. (“A claim that the verdict was against the weight of the evidence shall be raised with the trial judge in a motion for a new trial”)
Probative value = weight

- **Relevance**
  - Change in probability

- **Probative value**
  - Strength-weight

- **Sufficiency**
  - Burden of persuasion
Measures of Probative Value Proposed in Legal Literature

- $P(H|E) - P(H)$
  - Cullison 1969; Gerjuoy 1977; Friedman 1986

- $P(E|H) / P(E|\neg H)$

- $P(E|H) - P(E|\neg H)$
  - Davis & Follette 2002
WoE in the Three Fields

How lawyers speak about WoE

How statisticians speak about WoE

How forensic scientists speak about WoE
Theories of Inference

A statistical problem
- Data (x) are observed
- Probability model $P_\theta(x) = f(x|\theta)$ for generating $x$ has unknown values for its parameters $\theta$
- Having observed $x$, what can we say about $\theta$?
Three Approaches to Inference

**Frequentist**
- Fisher, Neymann, Pearson

**Likelihoodist**
- Barnard 1949; Edwards 1972; Royall 1997

**Bayesian**
- Bayes 1764
Frequentist

N-P hypothesis tests yield a decision

- A binary measure of the weight or strength of the evidence?

P-values indicate how surprising the evidence is under $H_0$: the smaller the p-value, the stronger the evidence against $H_0$.

- Characterizations of WoE: significant, highly significant, etc.

Example

- A fair and a biased coin: $H_0$ ($\theta = \frac{1}{2}$) and $H_1$ ($\theta = 1$)
- Data: 5 heads on 5 tosses
- $P = (1/2)^5 = 1/32 = 0.03$
Likelihood Defined

Probability distribution $P_\theta(x) = f(x|\theta)$

- $x$ varies, $\theta$ is fixed

Likelihood function $L(\theta) \propto f(x|\theta)$

- $\theta$ varies, $x$ is fixed

Example

- A fair and a biased coin: $H_0$ ($\theta = \frac{1}{2}$) and $H_1$ ($\theta = 1$)
- Data: 5 heads on 5 tosses
- Likelihood ratio $LR = \frac{L(H_1)}{L(H_0)} = \frac{k P(x|H_1)}{k P(x|H_0)} = \frac{1}{(\frac{1}{2})^5} = 32$
Three Approaches to Inference

Frequentist
- Fisher, Neymann, Pearson

Likelihoodist
- Barnard 1949; Edwards 1972; Royall 1997

Bayesian
- Bayes 1764
Likelihoodism  (Edwards 1972)

Likelihood Principle

- All the information which the data provide concerning the relative merits of the hypotheses is contained in the likelihood ratio of those hypotheses on the data.

Law of Likelihood

- Within the framework of a statistical model, a particular set of data supports one statistical hypothesis better than another if the likelihood of the first hypothesis, on the data, exceeds the likelihood of the second hypothesis.

Support function $S(\theta)$

- $\ln LR$
Three Approaches to Inference

Frequentist
- Fisher, Neymann, Pearson

Likelihoodist
- Barnard 1949;
  Edwards 1972;
  Royall 1997

Bayesian
- Bayes 1764
Bayesian inference

\[ f(\theta|x) = a \cdot L(\theta) \cdot f(\theta) \]

where \(1/a = \int L(\theta)f(\theta)d\theta\)
Bayes Rule for Binary $\theta$

$$
\text{Odds}(H_1|E) = \frac{P(E|H_1)}{P(E|H_0)} \cdot \text{Odds}(H_1)
$$

$\text{Log odds}(H_1|E) = \text{Log} \ \frac{P(E|H_1)}{P(E|H_0)} + \text{Log odds}(H_1)$

- posterior odds on $H_1$
- Bayes factor for $H_1$
- prior odds on $H$
- posterior log-odds
- WoE (Good 1991)
- prior log-odds
Verbal Tags for BFs

<table>
<thead>
<tr>
<th>Log-BF</th>
<th>BF</th>
<th>Verbal tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 to ½</td>
<td>1 to 3.16</td>
<td>barely worth mentioning</td>
</tr>
<tr>
<td>½ to 1</td>
<td>3.16 to 10</td>
<td>substantial</td>
</tr>
<tr>
<td>1 to 1½</td>
<td>10 to 31.6</td>
<td>strong</td>
</tr>
<tr>
<td>1½ to 2</td>
<td>31.6 to 100</td>
<td>very strong</td>
</tr>
<tr>
<td>&gt; 2</td>
<td>&gt; 100</td>
<td>decisive</td>
</tr>
</tbody>
</table>

- Jeffreys 1961
Example: The 2 Coins

\[ BF = LR = \frac{P(E|H_1)}{P(E|H_0)} = 1 \left(\frac{1}{2}\right)^5 = 32 \]

\[ WoE = \log(BF) = \log \frac{P(E|H_1)}{P(E|H_0)} = 1.51 \]

<table>
<thead>
<tr>
<th>BF (LR)</th>
<th>WoE (log-LR)</th>
<th>Jeffreys tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>1.51</td>
<td>very strong</td>
</tr>
</tbody>
</table>
WoE in the Three Fields

- How lawyers speak about WoE
- How statisticians speak about WoE
- How forensic scientists speak about WoE
How FSWs Describe PV in Court

- They do not
  - Features only
  - Match

- They go directly to hypotheses
  - Categorical conclusion
  - Posterior probability

- They state values
  - Freq or P quant
  - Freq or P qual
  - LR (quant & qual)
How FSWs Describe PV in Court

They do not

- Features only
- Match
[A hair examiner] displayed an enlarged photograph of one of the defendant's hairs and one of the hairs recovered from the victim's clothing as they appeared side-by-side under the comparison microscope. [He] explained to the jurors how the hairs were similar and what particular features of the hairs were visible. He also drew a diagram of a hair on a courtroom blackboard for the jurors. The jurors were free to make their own determinations as to the weight they would accord the expert's testimony in the light of the photograph and their own powers of observation and comparison.
Match testimony

tested that some of those hairs were consistent, meaning had the same characteristics, with known hair samples provided by [the defendant] and some of those hairs were consistent with hair samples from the victim . . . .”


tested that . . . a piece of cord taken from the scene of the crime [and] a piece of cord taken from the hood of a jacket ‘matched each other in component structure, . . . were similar and could have . . . originated from the same jacket.’

• State v. Gomes, 881 A. 2d 97 (R.I. 2005)

not error to admit “testimony that [defendant] could not be excluded as the source of the DNA obtained from the sneakers [even without] testimony explaining the statistical relevance of the nonexclusion result, such as the percentage of the population that could be excluded.”

How FSWs Describe PV in Court

They go directly to hypotheses

- **Categorical conclusion**
- **Posterior probability**
Categorical Conclusions

[A] ballistics expert, testified that he examined and compared the single shell casing found at the scene with the shell casings from the test firing of the gun found in the backpack. Walsh gave an opinion that to a ‘reasonable degree of certainty in the ballistics community’ the spent shell casing from the scene and the shell casings from the test firing were fired from the same weapon.


For each elemental ratio, compare the average ratio for the questioned specimen to the average ratio for the known specimens ±3σ. This range corresponds to 99.7% of a normally distributed population.

• ASTM 2926-13 (μ-XRF spectrometry)
The blood genetic marker tests ... registered a composite 99.99% probability that he is the biological father of Baby C, ... the chance of someone else ... is one in ten thousand.


Christina Buettner from the Wyoming State Crime Lab first testified “the probability of paternity” is “99.9999998638” that Mr. Snyder is the father of JL's baby.

How FSWs Describe PV in Court

They state values

- Freq or P quant
- Freq or P qual
- LR (quant & qual)
Frequency or P (Qualitative)

“an uncommon type of glass”


“no doubt that” the impressions of “two very experienced forensic scientists” that 20 DNA alleles consistent with the defendant’s full genotype were “rare” or at least “somewhat unusual” was “of assistance to a jury”

Only 3.8 out of 100 samples could have the same physical properties, based upon the refractive index test alone, which was performed.


1 in many quadrillions, quintillions, sextillions, or septillions

Mixtures: 1/1 to 1/10, 1/80, 1/300, 1/500, 1/3,000, 1/8,000, 1/9,000, 1/15,000, 1/35,000, 1/120,000, and 1/180,000

- DNA cases
Likelihood Ratios

DNA mixture cases (and some others)

Paternity, siblingship, avuncular index
### Qualitative LRs

#### Evett 1991

<table>
<thead>
<tr>
<th>Log LR</th>
<th>LR</th>
<th>Verbal Tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 to ½</td>
<td>1 to 33</td>
<td>weak</td>
</tr>
<tr>
<td>½ to 2</td>
<td>33 to 100</td>
<td>fair</td>
</tr>
<tr>
<td>2 to 2½</td>
<td>100 to 330</td>
<td>good</td>
</tr>
<tr>
<td>2½ to 3</td>
<td>330 to 1000</td>
<td>strong</td>
</tr>
<tr>
<td>&gt;3</td>
<td>&gt;1000</td>
<td>very strong</td>
</tr>
</tbody>
</table>

#### ENFSI 2015

<table>
<thead>
<tr>
<th>Log LR</th>
<th>Verbal Tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>no support</td>
</tr>
<tr>
<td>0.3 to 1</td>
<td>weak</td>
</tr>
<tr>
<td>2 to 3</td>
<td>moderate</td>
</tr>
<tr>
<td>2 to 3</td>
<td>strong</td>
</tr>
<tr>
<td>4 to 6</td>
<td>very strong</td>
</tr>
<tr>
<td>&gt;6</td>
<td>extremely strong</td>
</tr>
</tbody>
</table>

Approved of in NRC 2009
WoE in the Three Fields

How lawyers speak about WoE

How statisticians speak about WoE

How forensic scientists speak about WoE
References

- Workshop bibliography +
- ---, Likelihoodism, Bayesianism, and a Pair of Shoes, 53 Jurimetrics J. 1-9 (2012)